RT. A Journal on Research Policy & Evaluation **1** (2013)
Submitted on 18 November 2013, accepted on 26 Novembre 2013, published on 28 November 2013

Doi: 10.13130/2282-5398/3378

What ranking journals has in common with astrology

Björn Brembs*

1. Introduction

As scientists, we all send our best work to *Science* or *Nature* – or at least we dream of one day making a discovery we deem worthy of sending there. So obvious does this hierarchy in our journal landscape appear to our intuition, that when erroneous or fraudulent work is published in 'high-ranking' journals, we immediately wonder how this could have happened. Isn't work published there the best there is? Vetted by professional editors before being sent out to the most critical and most competent experts this planet has to offer? How could our system fail us so badly? We are used to boring, ill-designed, even flawed research in the 'low-ranking' journals where we publish. Surely, these incidents in the 'top' journals are few and far between?

It may come as a surprise to many scientists that the data speak a different language. They indicate that perhaps erroneous and fraudulent work is more common in 'top' journals than anywhere else (Brembs et al., 2013). There is direct evidence that the methodology of the research published in these journals is at least not superior, perhaps even inferior to work published elsewhere (Brembs et al., 2013) There is some indirect evidence that the error-detection rate may be slightly higher in 'top' journals, compared to other journals (Brembs et al., 2013). Neither data alone are sufficient to explain why high-ranking journals retract so many more studies than lower-ranking journals, but together they raise a disturbing suspicion: attention to top journals shapes the content of our journals more than scientific rigor. The attention being paid to publications in high-ranking journals not only entices scientists to send their best work to these journals, it also attracts fraudsters as well as unexpected and eye-catching results, which all too often prove literally too good to be true (Steen, 2011a, 2011b; Fang and Casadevall, 2011; Cokol et al., 2007; Hamilton, 2011; Fang et al., 2012; Wager and Williams, 2011). A conservative interpretation of the currently available data suggests that the attraction for truly groundbreaking, solid research just barely cancels out the attraction for unreliable or fraudulent work. A less conservative approach suggests that the solid research is losing.

How can the data be so at odds with our intuition? Of course, the research providing these data may itself be flawed. Selection bias, methodological errors and field-specific distortions may be dominating the studies currently available. However, in the absence of any evidence of such

* University of Regensburg, bjoern@brembs.net

systematic flaws spreading over many studies covering the last two decades, perhaps we at least ought to consider that our intuitions may be failing us. Perhaps journal rank's allure is no better than a horoscope – it only works its magic when you read your own horoscope and marvel at how well it describes you and how well its predictions have been met in the past. If the currently available data are taken at face value, they indicate that confirmation bias and selective attention on the part of scientists may be the only factors keeping *Nature* and *Science* at the top. At least, there is nothing in the current literature that can be taken as an evidence-based justification for their standing.

The current official justification for the rank of a journal, Thomson Reuters' Journal Impact Factor (JIF), is probably the worst possible attempt at such a ranking. In contrast to widespread belief, the JIF only appears to be calculated on hard citation data. It is by now thoroughly documented that it is only loosely based on a select subset of some citations (Editorial, 2005; Garfield, 1999; Adam, 2002; Moed and Van Leeuwen, 1995; Moed and van Leeuwen, 1996; Hernán, 2009; Baylis et al., 1999; Brembs et al., 2013). First, publishers negotiate with Thomson Reuters which articles are being used for the calculation (Baylis et al., 1999; The PLoS Medicine Editors, 2006; Brembs et al., 2013) , leading to documented shifts in a journal's JIF of an order of magnitude or more after such negotiations e.g., (Baylis et al., 1999). Second, when these negotiated data are subjected to independent verification, they fail to be reproduced, suggesting that after the negotiations with the publishers, Thomson Reuters is altering JIF values in a non-reproducible way before they are published (Rossner et al., 2007). Third, even if the JIF data could be relied upon, the statistical concept behind the JIF violates even the most basic mathematical rules: the JIF is calculated as a mean citation rate, when it has been known for decades that citation data are nowhere near normally distributed (Rossner et al., 2007; Seglen, 1992, 1997; Kravitz and Baker, 2011; Editorial, 2005; Chow et al., 2007; Weale et al., 2004; Taylor et al., 2008; Adler et al., 2008).

While there are many more flaws to be listed (e.g., (Vanclay, 2011), these three fundamental flaws alone help explain why the JIF has about as much predictive power as a horoscope. The actual citations an article receives are only very weakly correlated with the JIF of the journal the article was published in (Chow et al., 2007; Kravitz and Baker, 2011; Singh et al., 2007; Starbuck, 2005; Lozano et al., 2012; Hegarty and Walton, 2012; Seglen, 1997; Callaham, 2002; Seglen, 1994; Finardi, 2013). Adding insult to injury, the data suggest that reading habits, rather than any content of the journals are the main explanation for this relationship (Lozano et al., 2012).

Thus, much like astrologists use unrelated factors such as the position of stars in the sky to judge and predict a person's qualities, scientists use unrelated factors, such as the container within which an article was published, to judge and predict the qualities of an article. As a scientist, one would hope that the scientific community would be less impervious to data than astrologists. If not, it probably deserves what may be happening on or before the year 2046. This is the year when one retraction will be issued for every new article published (i.e., a 100% retraction rate), provided the current exponential trend in the rise of retractions (based on PubMed data) continues unabated (Fig. 1a (Brembs et al., 2013).

What could an astrology-like method of evaluating scientific publications have to do with an exponentially increasing retraction rate? By and large, currently tenured faculty have received tenure by publishing in highly ranked journals. In times of fewer and fewer tenured positions with more and more STEM graduates, today's faculty train their students, consciously or subconsciously, that publication in top journals is a condition for remaining and getting ahead in science. The data suggests that rather than scientific quality, the potential to market one's research for a general audience is more of a determining factor for obtaining publications in high-ranking journals. With selection for tenured faculty hence focusing more on their marketing than their scientific skills and the second generation of such scientists now starting their own labs, an exponential rise in the unreliability of scientific publications is –all else being equal- precisely what one would expect. It has to be emphasized that at this point this is mere evidence-based speculation – there is no direct evidence in support of this interpretation. Notwithstanding, the hypothesis is consistent with the currently available data.

2. Why the scientific infrastructure needs reform

The potential catastrophe indicated by the data is not the only reason to argue for publication reform. Even if later studies were to prove a majority of the doom-and-gloom predicting studies wrong, the dysfunctionality of our current communication system is cause for reform by itself. Not only do we rely on 20th century citation technology for linking our publications together, we also cannot access all of it, we mostly don't know who is reading our papers and we are barred from using modern information technology to assist us in selecting what to read. To make matters even worse, we are overpaying publishers for their services, by some accounts by as much as over 90%, compared to alternative systems offering similar or even superior services (see below).

Thus, publication reform is badly needed, even if there wasn't a major crisis looming. However, with the computer age, publications are increasingly being complemented by two other fruits of our intellectual labor: data and software. If the dysfunctional state of our literature can already be described as abysmal, our current digital infrastructure for data and software is even worse. Field-specific databases exist largely as grant-funded projects and are hence not on sustainable financial footing (Baker, 2012; Editorial, 2009; Merali and Giles, 2005; Editorial, 2012). For the last two decades, many, even major databases have been closed or have faced severe funding crises. The recent shutdown of the US government has brought the fickle nature of even government-funded sequence databases such as those hosted by the NCBI into everybody's awareness. Rounding off the trio of infrastructure horrors is scientific software for which there is no academic infrastructure whatsoever (Ince et al., 2012; Peng, 2011; Joppa et al., 2013; Mesirov, 2010).

3. How scientific institutions can avert the looming crisis by themselves

However, desperation is not the right response when faced with such challenges. In particular, the crisis science faces can be solved by the scientific institutions themselves, without the assistance of any outside entities. The solution lies in the combination of the expert know-how residing in the computing centers and libraries of our institutions with the vast subscription funds currently flowing into the pockets of the shareholders of a few international publishing houses. Current digital technology for an infrastructure serving all our intellectual fruits, software, data and their text summaries, is cheap and readily available. The required know-how already exists within our institutions and the funds are also available, albeit still locked in publisher contracts. The only factor missing for an infrastructure that promises to solve all the above problems is international coordination, in order to transition away from a legacy system that is not even close to serve even the most basic needs of the scientific community.

Libraries already digitally archive and make accessible many journals, on top of our theses and other works. The SciELO platform (http://scielo.org) in developing countries has been demonstrating for 15 years now that fully accessible, peer-reviewed publishing can be established at US$90 per article (Packer, 2010), rather than the US$4800 the developed countries are currently paying the publishers, e.g. (Van Noorden, 2013). Scientists have been running databases for research data for decades and hosting version-controlled, multi-user software repositories is not rocket science any more. Thus, we already have the expertise and the funds in our institutions to expand the existing literature infrastructure, complement it by migrating external databases in-house and extend it by implementing a software infrastructure.

This infrastructure would save money, save lives, save scientists time and allow for the development of an evidence-based reputation system to boot. There simply is no reason to delay reform any further.

References

Adam, David (2002), "The counting house" *Nature* 415, 726–9.

Adler, Robert, Ewing, John, and Taylor, Peter (2008). Joint Committee on Quantitative Assessment of Research: *Citation Statistics* (A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathemat.

Baker, Monia (2012). Databases fight funding cuts. *Nature* 489, 19–19.

Baylis, Matthew., Gravenor, Michael, and Kao, Roland (1999). "Sprucing up one's impact factor". *Nature* 401, 322.

Brembs, Björn, Button, Katherine, and Munafò, Marcus (2013). "Deep impact: unintended consequences of journal rank", *Frontiers in Human Neuroscience* 7, 291.

Callaham, Michael, Wears Robert, Weber Ellen (2002), "Journal Prestige, Publication Bias, and Other Characteristics Associated With Citation of Published Studies in Peer-Reviewed Journals", *JAMA Journal American Medical Association* 287, 2847–2850.

Chow, Chee W., Haddad, Kamal, Singh, Gangaram, and Wu, Anne (2007) "On Using Journal Rank to Proxy for an Article ' s Contribution or Value", *Issues in Accounting Education* 22, 411–427.

Cokol, Murat, Iossifov, Ivan, Rodriguez-Esteban, Raul, and Rzhetsky, Andrey (2007), "How many scientific papers should be retracted?" *EMBO Reports* 8, 422–3.

Editorial (2009), "Access denied?" *Nature* 462, 252.

Editorial (2005), "Not-so-deep impact" *Nature* 435, 1003–1004.

Editorial (2012), "The BMRB matters" *Nature Structural and Molecular Biology* 19, 853–853.

Fang, Ferric C. and Casadevall, Arturo (2011), "Retracted science and the retraction index" *Infection and Immunity* 79, 3855–9.

Fang, Ferric C., Steen, Grant R., and Casadevall, Arturo (2012). "Misconduct accounts for the majority of retracted scientific publications" *PNAS U. S. A.* 109, 17028–33.

Finardi, Ugo (2013), "Correlation between Journal Impact Factor and Citation Performance: An experimental study". *Journal of Informetrics* 7, 357–370.

Garfield, Eugene (1999). "Journal impact factor: a brief review" *CMAJ* 161, 979–80.

Hamilton, Jon (2011). "Debunked Science: Studies Take Heat In 2011". *NPR*. Available at: http://www.npr.org/2011/12/29/144431640/debunked-science-studies-take-heat-in-2011 [Accessed March 8, 2012].

Hegarty, Peter, and Walton, Zoe (2012) "The Consequences of Predicting Scientific Impact in Psychology Using Journal Impact Factors" *Perspectives on Psychological Science* 7, 72–78.

Hernán, Miguel A. (2009), "Impact factor: a call to reason" *Epidemiology* 20, 317–8; discussion 319–20.

Ince, Darrell C., Hatton, Leslie, and Graham-Cumming, John (2012) "The case for open computer programs", *Nature* 482, 485–8.

Joppa, Lucas N., McInerny, Greg, Harper, Richard, Salido, Lara, Takeda, Kenij, O'Hara, Kenton, Gavaghan, David, and Emmott, Stephen (2013), "Computational science. Troubling trends in scientific software use", *Science* 340, 814–5.

Kravitz, Dwight J., and Baker, Chris I. (2011), "Toward a new model of scientific publishing: discussion and a proposal" *Frontiers in Computational Neuroscience* 5, 55.

Lozano, George A., Larivière, Vincent, and Gingras, Yves (2012). "The weakening relationship between the impact factor and papers' citations in the digital age" *JASIST* 63, 2140–2145.

Merali, Zeeya, and Giles, Jim (2005), "Databases in peril" *Nature* 435, 1010–1.

Mesirov, Jill P. (2010), "Computer science. Accessible reproducible research", *Science* 327, 415–6.

Moed, Henk F., and van Leeuwen, Ted N. (1996), "Impact factors can mislead" *Nature* 381, 186.

Moed, Henk F., and Van Leeuwen, Ted N. (1995),"Improving the accuracy of institute for scientific information's journal impact factors" *JASIST*. 46, 461–467.

Van Noorden, Richard (2013)"Open access: The true cost of science publishing" *Nature* 495, 426–9.

Packer, Abel L. (2010), "The SciELO Open Access: A Gold Way from the South" *Canadian Journal of Higher Education* 39, 111–126.

Peng, Roger D. (2011), "Reproducible research in computational science" *Science* 334, 1226–7.

Rossner, Mike, Van Epps, Heather, and Hill, Emma (2007). "Show me the data". *Journal of Cellular Biology* 179, 1091–1092.

Seglen, Per O. (1994) "Causal relationship between article citedness and journal impact" *JASIS* 45, 1–11.

Seglen, Per O. (1992), "The skewness of science" *JASIS* 43, 628–638.

Seglen, Per O. (1997), "Why the impact factor of journals should not be used for evaluating research" *BMJ* 314.

Singh, Gangaram, Haddad, Kamal M., and Chow, Chee W. (2007), "Are Articles in "Top" Management Journals Necessarily of Higher Quality?" *Journal of Management Inquiry* 16, 319–331.

Starbuck, William H. (2005), "How Much Better Are the Most-Prestigious Journals? The Statistics of Academic Publication" *Organization Science* 16, 180–200.

Steen, Grant R., (2011a), "Retractions in the scientific literature: do authors deliberately commit research fraud?" *Journal of Medical Ethics* 37, 113–7.

Steen, Grant R., (2011b), "Retractions in the scientific literature: is the incidence of research fraud increasing?" *Journal of Medical Ethics* 37, 249–53.

Taylor, Michael, Perakakis, Pandelis, and Trachana, Varvara (2008), "The siege of science" *Ethics in Science and Environmental. Politics* 8, 17–40.

The PLoS Medicine Editors (2006) "The impact factor game. It is time to find a better way to assess the scientific literature" *PLoS Med.* 3, e291.

Vanclay, Jerome K. (2011). "Impact factor: outdated artefact or stepping-stone to journal certification?" *Scientometrics* 92, 211–238.

Wager, Elizabeth, and Williams, Peter (2011), "Why and how do journals retract articles? An analysis of Medline retractions 1988-2008" *Journal of Medical Ethics* 37, 567–70.

Weale, Andy R., Bailey, Mick, and Lear, Paul A. (2004), "The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor" *BMC Medical Research Methodology* 4, 14.